Assessment and enhancement of MERRA land surface hydrology estimates

Rolf H. Reichle[1], Randal D. Koster[1], Gabriëlle J. M. De Lannoy[1,2,3], Barton A. Forman[1,4],

Qing Liu[1,5], Sarith P. P. Mahanama[1,5], Ally Touré[1,2]

[1]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, USA

[2]Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, MD, USA

[3]Laboratory of Hydrology and Water Management, Ghent University, Ghent, Belgium

[4]Oak Ridge Associated Universities, Oak Ridge, Tennessee, USA

[5]Science Applications International Corporation, Beltsville, Maryland, USA

Corresponding author address

Rolf H. Reichle
NASA Goddard Space Flight Center
Mail Code 610.1
8800 Greenbelt Road
Greenbelt, MD 20771
USA

Tel.:   301-614-5693
FAX: 301-614-6246
Email: rolf.reichle@nasa.gov

34 **Abstract**

35 The Modern-Era Retrospective analysis for Research and Applications (MERRA) is a state-of-

36 the-art reanalysis that provides, in addition to atmospheric fields, global estimates of soil

37 moisture, latent heat flux, snow, and runoff for 1979-present. This study introduces a

38 supplemental and improved set of land surface hydrological fields ("MERRA-Land") generated

39 by re-running a revised version of the land component of the MERRA system. Specifically, the

40 MERRA-Land estimates benefit from corrections to the precipitation forcing with the Global

41 Precipitation Climatology Project pentad product (version 2.1) and from revised parameter

42 values in the rainfall interception model, changes that effectively correct for known limitations in

43 the MERRA surface meteorological forcings. The skill (defined as the correlation coefficient of

44 the anomaly time series) in land surface hydrological fields from MERRA and MERRA-Land is

45 assessed here against observations and compared to the skill of the state-of-the-art ERA-Interim

46 (ERA-I) reanalysis. MERRA-Land and ERA-I root zone soil moisture skills (against in situ

47 observations at 85 US stations) are comparable and significantly greater than that of MERRA.

48 Throughout the northern hemisphere, MERRA and MERRA-Land agree reasonably well with in

49 situ snow depth measurements (from 583 stations) and with snow water equivalent from an

50 independent analysis. Runoff skill (against naturalized stream flow observations from 18 US

51 basins) of MERRA and MERRA-Land is typically higher than that of ERA-I. With a few

52 exceptions, the MERRA-Land data appear more accurate than the original MERRA estimates

53 and are thus recommended for those interested in using MERRA output for land surface

54 hydrological studies.

1.   Introduction

The Modern-Era Retrospective analysis for Research and Applications (MERRA; Rienecker et al. 2011) is a recent addition to the suite of global, long-term reanalysis products that are based on the assimilation of in situ and remote sensing observations into numerical models of the global atmosphere and land surface (Kalnay et al. 1996; Kanamitsu et al. 2002; Uppala et al. 2005; Onogi et al. 2007; Dee et al. 2011; Saha et al. 2010). Besides estimates of atmospheric conditions, reanalysis products also provide estimates of land surface fields, including surface meteorological forcing data (such as precipitation, radiation, air temperature, and humidity) as well as land surface states and fluxes (such as soil moisture, snow, and runoff). Reanalysis estimates can be used for a large variety of research and applications, for example the generation of enhanced land surface meteorological data sets (Berg et al. 2005; Guo et al. 2006; Sheffield et al. 2006), the study of the land surface water budget, including streamflow, droughts, soil moisture, and snow processes (Dai and Trenberth 2002; Su and Lettenmaier 2009; Sheffield and Wood 2009; Burke et al. 2010; Brown et al. 2010), the estimation of the land carbon budget (Zhao et al. 2006; Yi et al. 2011), and, possibly, the calibration and verification of seasonal climate forecasting systems (Saha et al. 2006) and the generation of climate data records (Thorne and Vose 2010; Dee et al. 2010).

The MERRA data products are available from 1979 to present at high spatial and temporal resolution and are based on the assimilation of a vast number of *atmospheric* observations. MERRA land surface estimates, however, utilize no directly assimilated land surface observations; they reflect instead the time integration of surface meteorological conditions (precipitation, radiation, wind speed, etc.) by the land model component of MERRA. Based on

2

77 the analyzed atmospheric state (including humidity and temperature profiles), MERRA

78 precipitation over land is generated by the Atmospheric General Circulation Model (AGCM)

79 during the Incremental Analysis Update segment (Rienecker et al. 2011) and is thus subject to

80 considerable errors that ultimately propagate into the land surface hydrological fields. Moreover,

81 errors in land surface estimates result from errors in the land surface model itself, including

82 imperfect representation of physical processes and uncertainties in the land model parameters.

83 Given knowledge of such errors, it is reasonable to attempt to mitigate their impacts through the

84 careful post-processing of MERRA output. Such post-processing, if done properly, could

85 produce a land surface dataset more useful and appropriate for hydrological analyses. Here, we

86 describe a particular post-processing of the MERRA land fields that involves the reintegration of

87 the land surface model with more realistic precipitation forcing and with a parameterization

88 change designed to counteract certain known problems with MERRA's diurnal rainfall and

89 radiation cycles. The resulting fields, along with the original MERRA land fields, are compared

90 extensively to observations; advantages of the post-processed dataset (hereinafter "MERRA-

91 Land") are highlighted.

92 We emphasize that these known problems are typical of global reanalysis data products. On

93 average, global precipitation from MERRA is no worse than estimates from other reanalysis

94 products (Bosilovich et al. 2011). There have been many similar efforts to improve global off-

95 line land surface simulations through corrected analysis or reanalysis forcing data (for example,

96 Dirmeyer and Tan 2001; Berg et al. 2005; Guo et al. 2006; Qian et al. 2006; Sheffield et al.

97 2006). Our paper focuses on the land surface hydrology estimates from MERRA and how they

3

98   can be improved through simple corrections to land model parameters and the precipitation

99   forcing.

100   The paper is organized as follows. Section 2 briefly describes the MERRA modeling system and

101   data product, along with the data used for its evaluation. Section 3 starts with a brief evaluation

102   of MERRA surface precipitation and radiation estimates and motivates the development of the

103   MERRA-Land product, which is described in detail thereafter. Section 4 evaluates MERRA and

104   MERRA-Land estimates of interception loss fraction, latent heat flux, soil moisture, runoff, and

105   snow. Additional discussion and conclusions follow in section 5. Appendix A details the skill

106   metric used herein.

107

2.   Data

a. The MERRA system and data product

MERRA is a reanalysis product generated by the NASA Global Modeling and Assimilation

Office (GMAO) using the Goddard Earth Observing System (GEOS) version 5.2.0 (Rienecker et

al. 2011; http://gmao.gsfc.nasa.gov/research/merra/). The system incorporates information from

in situ and remote sensing observations of the atmosphere, including many modern satellite

observations such as Atmospheric Infrared Sounder (AIRS) radiances and scatterometer-based

wind retrievals. These observations are assimilated into the GEOS-5 AGCM using the National

Centers for Environmental Prediction Gridpoint Statistical Interpolation assimilation package.

MERRA, however, does not include a land surface analysis. MERRA covers the period from

1979 to the present and continues to be updated with latency on the order of weeks. MERRA

estimates of surface meteorological and land surface fields are available at hourly time steps and

at $1/2° \times 2/3°$ resolution in latitude and longitude, respectively.

The GEOS-5 AGCM includes a set of state-of-the-art physics packages, along with the

innovative GEOS-5 Catchment land surface model (hereinafter Catchment model; Koster et al.

2000; Ducharne et al. 2000). The model is designed to improve the treatment of land surface

hydrological processes through explicit modeling of sub-grid scale soil moisture variability and

its effect on runoff and evaporation. The basic computational unit of the model is the

hydrological catchment (or watershed), with boundaries defined by topography (see below).

Within each element, the vertical profile of soil moisture is given by the equilibrium soil

moisture profile and the deviations from the equilibrium profile (described by variables in a 0-2

129   cm surface layer and in a "root zone" layer that extends from the surface to a depth z_R, with 75

130   cm $\leq z_R \leq$ 100 cm depending on local soil conditions). The spatial variability of soil moisture is

131   diagnosed at each time step from the bulk water prognostic variables and the statistics of the

132   catchment topography. The soil and vegetation parameters used in the Catchment model are from

133   the NASA GEOS-5 global modeling system (Rienecker et al. 2011). The Catchment model also

134   includes a state-of-the-art snow model (Stieglitz et al. 2001); in each watershed, the evolution of

135   snow water equivalent (SWE), snow depth, and snow heat content in response to surface

136   meteorological conditions and snow compaction is modeled using three layers. The time step for

137   the land model integration is 20 min.

138   The Catchment model's computations are performed at a higher spatial resolution than those of

139   the atmosphere. The basic land surface element, or "tile", is a topographically determined

140   hydrological catchment; catchments that straddle AGCM grid cells are subdivided by the grid

141   boundary into smaller tiles. Although standard MERRA output is available only on the 1/2° ×

142   2/3° grid, higher-resolution tile-based land surface fields are generated (but not saved) as part of

143   the MERRA data production. For MERRA, the Catchment model uses 157,051 land tiles with a

144   mean (median) area of 828 km^2 (524 km^2), resulting in an average resolution of about 25 km.

145   For this study, we "replayed" the MERRA land surface component by forcing the Catchment

146   model off-line (that is, not coupled to the atmospheric model) after interpolation of the hourly

147   land surface meteorological fields from the standard MERRA output to the 20 minute Catchment

148   model time step. The replay configuration produces output that is only marginally different from

149   the original MERRA land surface fields, and it serves two important purposes. First, it allows us

150   to conduct the skill assessment using the higher-resolution tile output and thereby lessen the

151   impact of the discrepancy between the horizontally distributed scale of the model-based

152   estimates and the point-scale of the validating in situ measurements. Second, the MERRA-Land

153   estimates (discussed below) are based on the off-line replay configuration by construction, and

154   thus comparing them to the MERRA estimates generated offline under replay mode allows a

155   more careful isolation of the impacts of the precipitation corrections and model parameter

156   revisions on the accuracy of the product.

157   *b. Evaluation data*

158   1) PRECIPITATION OBSERVATIONS

159   We use the Global Precipitation Climatology Project (GPCP) precipitation pentad (5-day)

160   product version 2.1 (Huffman et al. 2009; Xie et al. 2003) to evaluate and correct the MERRA

161   precipitation estimates. The GPCP data are available as pentad averages from 1979 to 2009 on a

162   2.5˚ x 2.5˚ global grid and are based on the merging of satellite measurements (infrared and

163   microwave) with global rain gauge observations from the Global Precipitation Climatology

164   Centre. Specifically, the GPCP pentad product is computed by adjusting the pentad estimates

165   from the NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP; Xie and

166   Arkin 1997; http://www.esrl.noaa.gov/psd/data/gridded/data.cmap.html) product to monthly

167   GPCP version 2.1 estimates. GPCP and CMAP estimates differ primarily in the input and

168   processing of the satellite observations and in the approach for combining the satellite and gauge

169   inputs.

170

171   2) SOIL MOISTURE OBSERVATIONS

172   In situ soil moisture observations from the United States Department of Agriculture Soil Climate

173   Analysis Network (SCAN, Schaefer et al. 2007, http://www.wcc.nrcs.usda.gov) are used to

174   assess skill. Hourly soil moisture measurements were taken with a device measuring the

175   dielectric constant of the soil (Stevens Water Hydra Probe sensors inserted horizontally at depths

176   of 5 cm, 10 cm, 20 cm, 50 cm and 100 cm wherever possible). There are a total of 125 SCAN

177   sites in the contiguous United States that provide some data between 1 January 2002 and 31 July

178   2009, the period considered here (Figure 1). For data from each SCAN site we applied extensive

179   quality control steps that included automatic detection of problematic observations and a visual

180   inspection of the time series. We excluded data that are obviously unrealistic (such as data

181   outside of the physical range, or data related to discontinuities in the time series that could not be

182   explained by physical processes). We also excluded soil moisture measurements that were taken

183   under frozen conditions (according to SCAN soil temperature measurements), or data affected by

184   inconsistencies that are most likely due to changes in sensor calibration or sensor installation.

185   After quality control of the hourly data, the SCAN observations were aggregated into pentad

186   averages. Because of the quality control and the data requirements for the anomaly computation

187   (Appendix A), only 98 SCAN sites could be used to assess the skill of surface soil moisture

188   estimates, and only 85 of the 98 sites could be used to assess the skill of root zone moisture

189   estimates. Liu et al. (2011) discuss the validity of using the single-profile (point-scale) SCAN

190   measurements to assess the skill of land model estimates of soil moisture that represent average

191   values across tiles or grid cells.

192

8

193   3) STREAMFLOW OBSERVATIONS

194   Streamflow gauge data for 18 basins in the United States, ranging in size from 1,900 km^2 to

195   1,400,000 km^2, were used to assess runoff estimates (Table 1; see Koster et al. 2010 and

196   Mahanama et al. 2011 for details). The streamflow data were naturalized to account for

197   anthropogenic impacts, including upstream regulation, water withdrawals, and evaporation from

198   reservoir surfaces. Note that some of the basins used by Mahanama et al. (2011) lack sufficient

199   observations during our study period and are thus not considered here.

200   4) SNOW OBSERVATIONS

201   World Meteorological Organization (WMO) snow depth measurements were obtained from the

202   National Climatic Data Center (Tedesco and Miller 2010; http://nsidc.org/data/nsidc-0450.html).

203   A total of 583 stations located in the northern hemisphere (mostly in Russia, Europe, and Alaska)

204   for the period October 2002 through August 2009 were used because they fulfilled the screening

205   criteria outlined in Appendix A. In addition, we used the snow depth product from the Canadian

206   Meteorological Centre (CMC) daily snow analysis (Brasnett 1999; Brown and Brasnett 2010).

207   The CMC product provides daily snow depth throughout the northern hemisphere at a horizontal

208   resolution of approximately 24 km for the period of March 1998 to the present. The CMC snow

209   analysis is based on optimal interpolation of in situ daily snow depth observations and aviation

210   reports with a first-guess field generated from a simple snow model driven by analyzed

211   temperatures and forecast precipitation from the Canadian forecast model (Brasnett 1999). The

212   CMC product is often considered the "best available" snow depth product for the northern

213   hemisphere and has been used for evaluating model output (e.g., Su et al. 2010). Finally, Sturm

9

214 et al. (2010) provide climatological snow density estimates as a function of snow depth, day of

215 year, and snow class (except for the "ephemeral" snow class; see their equation 6). Using the

216 snow class map shown in Sturm et al. (1995) we obtained SWE estimates by multiplying the

217 CMC snow depths with the Sturm et al. (2010) snow densities for subsequent comparison against

218 SWE estimates from MERRA and MERRA-Land.

219 5) ERA-INTERIM

220 Whenever possible, we compare the skill of MERRA and MERRA-Land to that of ERA-Interim

221 (ERA-I), the most recent reanalysis product of the European Centre for Medium-Range Weather

222 Forecasts (Dee et al. 2011; http://www.ecmwf.int/research/era). Here, we use the daily ERA-I

223 data product that is publicly available at 1.5° resolution from 1989 to present (updated with

224 about two months latency). Soil moisture in ERA-I is modeled in four layers (0-7 cm, 7-28 cm,

225 28-100 cm, and 100-289 cm) and updated in response to screen-level (2 m) observations of air

226 temperature and humidity. This soil moisture analysis, however, is designed to improve the

227 turbulent surface flux estimates and subsequent atmospheric forecasts and provides no clear

228 benefit to soil moisture estimates (Drusch and Viterbo 2007). ERA-I also includes a snow

229 analysis based on in situ snow depth and satellite snow cover observations (Drusch et al. 2004).

230 The structure functions used in the ERA-I snow depth analysis differ from those used in the

231 CMC product. Because of recently discovered problems in the ECMWF system, the CMC

232 structure functions have been adopted in the latest version of the ECMWF operational system

233 (De Rosnay, ECMWF, personal communication, November 2010). Szczypta et al. (2011)

234 provide a detailed assessment over France of surface meteorological forcing data from ERA-I

235   (with and without corrections to monthly GPCP v2.1 precipitation estimates) and find that the

236   precipitation corrections lead to improved root zone soil moisture estimates.

237

3. Motivation for and construction of MERRA-Land

a. Motivation for a revised product

Precipitation is by far the most important driver of a land surface hydrological simulation; hence precipitation error will have an overwhelming impact on the accuracy of simulated hydrological fields regardless of the accuracy of the other forcings or the realism of the underlying land model. Although the spatial distribution of the MERRA mean annual precipitation is quite good compared to that of other reanalysis products (Bosilovich et al. 2011, see their figure 3), two correctable deficiencies associated with MERRA's precipitation forcing motivate our construction here of a revised land product: (1) inaccuracies in the climatological and synoptic variability of the precipitation forcing, and (2) inaccuracies in the intensity and diurnal cycle of this forcing.

1) LONG-TERM PRECIPITATION TOTALS

The precipitation estimates generated by MERRA do not benefit from the assimilation of surface rain gauge data. While they do benefit from the assimilation of water vapor, wind fields, and other atmospheric quantities (Rienecker et al. 2011), the onset, intensity, and cessation of any rainfall event is chiefly controlled by the model's precipitation parameterizations. (The assimilation in MERRA of satellite rain rate retrievals over the ocean has a negligible impact on the system over land.) As a result, MERRA precipitation fields show some inaccuracies relative to established, observations-based datasets, particularly over land, as will be shown next.

12

257   Figure 2a shows the mean annual precipitation for the period 1981-2008 from MERRA, and

258   Figure 2b shows the corresponding observations-based estimates from GPCP (section 2b).

259   MERRA and GPCP both have a global mean over land of around 2.3 mm d^{-1} for 1981-2008 (see

260   Bosilovich et al. 2011 for a discussion of the global water budget and trends of MERRA and

261   other reanalysis products). To first order, the precipitation fields look similar, with MERRA

262   locating deserts and rainy areas in the proper places and assigning, in most regions,

263   approximately the correct magnitudes to the mean annual precipitation rates. The MERRA

264   product, however, differs from the GPCP reference, as revealed by the difference map in Figure

265   2c. MERRA mean annual precipitation rates are biased low in much of South America and

266   central Africa and biased high in Southeast Asia, in Indonesia, and along the tropical South

267   American and African coasts. Smaller but still significant biases appear across much of the

268   globe. Note, however, that uncertainty in the GPCP precipitation estimates themselves, a strong

269   function of rain gauge density, varies significantly across the globe (Adler et al. 2003).

270   Figure 2d shows the difference field (MERRA minus GPCP) for a single representative month

271   (August 1994). Relative to those found for the long-term mean, the errors for this month are

272   reduced in parts of South America but are more often magnified, with values exceeding 1 mm d^{-1}

273   in many mid-latitude regions. Such errors will have a first order impact on the simulated land

274   surface hydrological variables. Our assumption in this paper is that "correcting" the MERRA

275   precipitation forcing so that it agrees with the GPCP data as much as possible should lead to

276   improved hydrological simulation.

277

278   2) INTENSITY AND DIURNAL CYCLE OF PRECIPITATION

279   Errors in the intensity and the diurnal cycle of precipitation are common in many atmospheric

280   modeling systems (Dai 2006). Unsurprisingly, MERRA also suffers from such deficiencies.

281   Figure 3 illustrates this with a representative example. The top panel shows MERRA time series

282   of solar radiation and precipitation for a 9-day summer period at a single grid cell near

283   Gainesville, Florida. The bottom panel shows the corresponding observations from a FLUXNET

284   site located within the MERRA grid cell. The MERRA time series differ from the FLUXNET

285   time series in at least three fundamental ways, each directly relevant to the simulation of

286   hydrological fluxes at the land surface. First, despite being similar in long-term average,

287   MERRA precipitation rates are less intense; relative to observations, MERRA rain tends to come

288   down as more of a long-lasting "drizzle". Second, the precipitation in MERRA tends to be

289   highest in the middle of the day, whereas the observations show frequent nighttime rain maxima.

290   Third, in the observations, a daytime precipitation event tends to reduce incoming solar radiation

291   substantially (e.g., on 21 June 2003), whereas in MERRA, the rain reduces the solar radiation by

292   only about half (16-20 June 2003) or sometimes hardly at all (23 June 2003).

293   The discrepancy between the distributed (grid cell) scale of the MERRA estimates (~50 km) and

294   the point scale of the in situ observations may be responsible for at least part of the rain intensity

295   and radiation differences shown in Figure 3. Nevertheless, regardless of their source, the three

296   features of MERRA rain and radiation behavior highlighted in the figure are commonplace for

297   MERRA summer precipitation and work together to confound the ability of MERRA to provide

298   adequate amounts of rainwater to the soil. Simply put, the drizzle of MERRA rainfall during

299   daylight hours – hours for which plenty of simulated solar radiation energy is available for

14

300   evaporation – leads to the immediate evaporation of much of this rainfall directly from droplets

301   sitting on the surface of the vegetation canopy (that is, directly from the land model's

302   interception reservoir). As a result, not enough of the water is allowed to drip down through the

303   canopy and ultimately infiltrate the soil or generate surface runoff. Relative to an off-line

304   simulation with the same land model but with more realistic forcing (e.g., along the lines of that

305   shown for the FLUXNET site), MERRA produces soil moistures that are too dry (section 4b),

306   with consequent impacts on the simulation of land surface hydrological fluxes.

307   *b. Construction of the MERRA-Land data product*

308   To mitigate the impacts of these problems, MERRA-Land estimates were generated by replaying

309   (that is, running off-line with prescribed and improved meteorological forcing) a revised version

310   of the land component of the MERRA system.

311   1) PRECIPITATION CORRECTIONS

312   For the new MERRA-Land product, all atmospheric forcing fields (including air temperature and

313   humidity, radiation, wind speed, and surface pressure) for the land surface model were taken

314   directly from hourly MERRA output, with one important exception: the MERRA precipitation

315   forcings were corrected towards gauge- and satellite-based observations using the GPCP version

316   2.1 pentad product (section 2b). Because of their coarse (pentad) time resolution, the GPCP data

317   themselves cannot be used to force the Catchment model. We therefore use the GPCP estimates

318   to construct a corrected version of the MERRA precipitation. The approach used here is similar

319   in concept to that applied in the Global Soil Wetness Project (Dirmeyer 2006) and other global

15

320 land modeling studies (Berg et al. 2005; Guo et al. 2006; Qian et al. 2006; Sheffield et al. 2006).

321 Based on results from these earlier studies, we recognize that corrections to surface radiation and

322 surface air temperature have a much smaller effect than precipitation corrections.   Such

323 additional forcing corrections could in any case lead to inconsistencies across the forcing fields

324 in cases where the observational data may be contradictory. Consequently, we restrict ourselves

325 here to correcting the precipitation forcing.

326 The corrected MERRA precipitation forcings were obtained as follows. First, the hourly

327 MERRA total precipitation was time-averaged and re-gridded to the scale of the correcting

328 GPCP dataset (that is, to pentad and 2.5˚ resolution). Next, for each pentad of each year and for

329 each 2.5˚ grid cell, a scaling factor was computed by determining the ratio of the GPCP estimate

330 to the standard MERRA data (that is, on the grid and at the time scale of the correcting

331 observations). Finally, these scaling factors were re-gridded back to the MERRA grid and

332 applied to the MERRA data – a scaling factor derived for a given grid cell and year/pentad was

333 applied to the MERRA precipitation rates (large-scale precipitation, convective precipitation, and

334 snowfall separately) in each of the 120 hourly time steps within that pentad. If for a given grid

335 cell the aggregated MERRA value was zero, the corresponding corrected MERRA precipitation

336 values were set to zero, even if the correcting observations indicated non-zero precipitation

337 (rather than distributing the observed precipitation across time steps in an ad hoc way) in order to

338 maintain consistency across the forcing variables (including surface radiation) to the fullest

339 extent possible. By construction, the corrected MERRA precipitation is nearly identical to the

340 GPCP estimates at the pentad and 2.5˚ resolution and is therefore not shown.

341   Because the GPCP product is based on precipitation observations from satellites and/or gauges

342   well beyond the data used in the MERRA atmospheric assimilation, we expect that the GPCP-

343   corrected MERRA precipitation forcing is more accurate than the standard MERRA precipitation

344   product. Note again, however, that the (hourly, $0.5°$) corrected precipitation dataset is a scaled

345   version of the MERRA precipitation forcing, rather than the original (pentad, $2.5°$) GPCP

346   dataset. The diurnal cycle, the frequency and relative intensity of rainfall events at the sub-

347   pentad scale, and the sub-$2.5°$ spatial variations are entirely based on MERRA estimates. While

348   Qian et al. (2006) discuss the possibility of also adjusting the diurnal cycle of the precipitation,

349   we choose here to impose the sub-pentad variations of the original MERRA precipitation in

350   order to maintain maximum consistency across the forcing variables (including surface

351   radiation). Finally, note again that the precipitation corrections are constructed separately for

352   each pentad of each year and thus go beyond a climatological adjustment.

353   2) CATCHMENT MODEL PARAMETER REVISIONS

354   The Catchment model version and model parameters used for MERRA-Land are identical to

355   those used for MERRA data production except for the changes to the interception and snow

356   parameters listed in Table 2. These changes bring the Catchment model used for MERRA-Land

357   up to date with the forthcoming version used in the GEOS-5 experimental NWP and seasonal

358   forecasting systems. Of particular relevance to the MERRA-Land product are the changes made

359   to the rainfall interception parameters FWETL and FWETC, changes that mitigate the impact of

360   the discrepancies outlined in Figure 3. These two parameters describe the fractional areas over

361   which large-scale and convective rainfall, respectively, are applied to the canopy interception

362   reservoir. In MERRA, large-scale rainfall is applied uniformly to the canopy (FWETL=1),

363   whereas the intensity of convective rainfall at a given time step is quintupled and applied to 1/5

364   of the area of the canopy (FWETC=0.2) – water is conserved, but the greater local depth allows

365   it (in principle) to overflow the interception reservoir and drip down to the surface more easily.

366   In MERRA-Land, this effect is heightened considerably: the intensity of either form of rainfall is

367   multiplied by fifty and applied to one-fiftieth of the canopy area (FWETL=FWETC=0.02). We

368   emphasize that this change is not meant to represent a realistic treatment of subgrid rainfall

369   variability. It is designed solely to circumvent known deficiencies in the atmospheric model's

370   representation of the intensity and diurnal cycle of rainfall and contemporaneous radiation

371   (Figure 3). The smaller fractional area of rainfall, while not realistic, does allow more of the

372   rainfall to drain through the canopy and reach the soil, leading to wetter soil and much more

373   sensible interception loss fractions (section 4a). It has no other impact on the simulation – in

374   particular, the prescribed one-fiftieth of the canopy area does not affect the partitioning of

375   throughfall into runoff and infiltration at the soil surface. Note that in other off-line applications

376   with the Catchment model, applications involving atmospheric forcing without the noted

377   problems, we can safely revert to the MERRA values for the two parameters.

378   Table 2 lists additional changes to the model parameters that bring the Catchment model up-to-

379   date with the forthcoming GEOS-5 version. The change in the capacity of the interception

380   reservoir (SATCAP) has an effect similar to that of the changes to FWETL and FWETC (albeit

381   much smaller, given the non-linear dynamics of the interception model). Moreover, changes

382   were made to the minimum SWE in the snow-covered area fraction (WEMIN) and the maximum

383   depth of the uppermost snow layer (DZ1MAX) to improve the modeled albedo and the stability

384   of the surface calculation when snow is present (not shown here). Because in the off-line replay

385   configuration of MERRA-Land the land fluxes do not feed back on the atmosphere, the snow

386   parameter changes lead to only minor differences between MERRA-Land and MERRA.

387

4. Results

In this section we evaluate land surface states and fluxes from MERRA and MERRA-Land against a variety of observations and independent model estimates. Our evaluation includes interception loss fraction and latent heat flux (section 4a), soil moisture (section 4b), runoff (section 4c), and snow (section 4d). Where appropriate, we also provide skill estimates for ERA-I (section 2b). We refer the reader to Yi et al. (2011) for a discussion of MERRA surface air temperature, vapor pressure deficit, and incident solar radiation. Yi et al. (2011) also provide additional analysis of MERRA surface soil moisture. Moreover, Decker et al. (2011) evaluate MERRA land surface forcings and fluxes against tower observations.

a. Interception loss fraction and latent heat flux

As discussed in section 2, the character of MERRA precipitation and radiation forcing is expected to have a detrimental effect on land surface hydrology. Perhaps the most striking effect is seen in the interception loss fraction I, defined as the fraction of incoming rainfall that is intercepted by the canopy and re-evaporated back to the atmosphere without ever infiltrating the soil or contributing to surface runoff. MERRA's long-term average I values, shown in Figure 4a, are greater than 0.24 almost everywhere, even in non-forested areas (for example, the US Great Plains) and occasionally in very sparsely vegetated areas (for example, the Sahara, and western and central Australia). In tropical rainforests, I values can exceed 0.5. Globally averaged, MERRA's interception loss fraction is I=0.31. These fractions are far in excess of published estimates, such as those of Miralles et al. (2010), shown in Figure 4d. The latter were derived by calibrating a global model of interception dynamics to a large number of in situ observations (see

20

409   references in (Miralles et al. 2010)). In their model, the largest I values, ranging from I=0.15 to

410   I=0.24, are found in the boreal forests of North America, Scandinavia, and Russia. Somewhat

411   smaller values of I=0.06 to I=0.15 are found in tropical rain forests (including Indonesia and the

412   Amazon and Congo basins) and mid-latitude forested regions (eastern United States, parts of

413   Europe). Globally averaged, Miralles et al. (2010) estimate I=0.06. For comparison, Sakaguchi

414   and Zeng (2009) report I=0.12 for the Community Land Model version 3.5.

415   Figure 4b shows the interception loss fractions for the revised Catchment model (Table 2, section

416   4b) when forced with MERRA surface meteorology. The revised interception parameters lead to

417   much more realistic I values, with a global average of I=0.07. In the boreal forest, the revised

418   Catchment model now underestimates the interception loss fraction (relative to the Miralles et al.

419   (2010) estimates), with values ranging between I=0.09 and I=0.21. In non-forested areas and

420   deserts, the interception loss fraction is now typically below I=0.09. However, errors in the

421   long-term climatology of MERRA precipitation still lead to I values greater than I=0.21 in the

422   Amazon and Congo basins. When the revised Catchment model is forced with the GPCP-

423   corrected precipitation (that is, MERRA-Land, shown in Figure 4c) the I values for these two

424   basins are reduced and agree well with the estimates from Miralles et al. (2010). Globally

425   averaged, the MERRA-Land interception loss fraction is I=0.07. The largest remaining

426   differences between I values from MERRA-Land and Miralles et al. (2010) are in the boreal

427   forests, where MERRA-Land estimates are lower.

428   The revised treatment of interception loss in MERRA-Land, combined with the GPCP-based

429   improvements in precipitation forcing, has impacts on other hydrological fields. Figure 5 shows

430   an example: MERRA estimates of latent heat flux (LH) for August 1994 are shown in Figure 5a,

21

431   and those for MERRA-Land are shown in Figure 5b. For reference, Figure 5c shows an estimate

432   based on 12 different products using a variety of data sources from remote sensing, flux tower

433   measurements, and land surface modeling (Jimenez et al., 2011). (MERRA is one of the 12

434   estimates in the multi-product average.) Overall, the three estimates agree reasonably well, with

435   global average LH values for this month of 58.0 W m^{-2} (MERRA), 55.4 W m^{-2} (MERRA-Land),

436   and 56.3 W m^{-2} (multi-product average). The three estimates also agree in the broad global

437   pattern of LH, with high values in the eastern US, the tropical rainforests, and south-east Asia.

438   Low values in the Southern Hemisphere are due to winter conditions in August.

439   One important difference between MERRA and the multi-product average LH, however, appears

440   in the Amazon basin. MERRA LH exhibits an extremely sharp north-south gradient, with values

441   quickly dropping from around 140 Wm^{-2} north of 5°S to less than 20 Wm^{-2} south of 8°S.   The

442   corresponding gradient in the multi-product average LH is much less steep, with values dropping

443   from 100 Wm^{-2} north of 7°S to 60 Wm^{-2} south of 15°S. Whereas MERRA could be considered

444   an outlier among the products evaluated by Jimenez et al. (2011), MERRA-Land is not – its LH

445   estimates lie within the range of estimates contributing to the multi-product average (not shown,

446   see their figure 6). Note that MERRA precipitation errors also exhibit a strong gradient along

447   5°S (Figure 2d). Additional analysis (not shown) indicates that the GPCP-based precipitation

448   corrections and the interception parameter revisions contribute about equally to the LH

449   improvements in MERRA-Land.

450

22

451   *b. Soil moisture*

452   The interception model revisions by themselves have important implications for soil moisture.

453   Again, the revised parameters were designed to let more of the incoming rainfall reach the soil

454   and thereby increase long-term soil moisture levels. This can be seen in Figure 6a, which shows

455   the difference between the 1981-2008 average root zone soil moisture from MERRA and from

456   the revised Catchment model (when forced with MERRA surface meteorology). Differences in

457   root zone soil moisture up to -0.05 m^3m^{-3} occur in the boreal forests, the south-eastern US, and

458   the Amazon and Congo basins, that is, in areas with generally moist climates and with the largest

459   changes in the interception loss fraction (Figure 4). As expected, soil moisture generated by the

460   revised Catchment model is always wetter than that of MERRA.

461   Figure 6b shows the combined impact of the GPCP-based precipitation corrections and the

462   Catchment model parameter revisions on long-term root zone soil moisture in MERRA-Land.

463   Unsurprisingly, the overall global pattern of the root zone soil moisture differences is dominated

464   by the differences in the precipitation forcing. Where MERRA precipitation is biased dry

465   against GPCP (Figure 2c), such as in much of South America and central Africa, MERRA-Land

466   root zone soil moisture is considerably higher because of the combined effect of higher

467   precipitation forcing and reduced interception (Figure 6b). Where MERRA precipitation is

468   biased wet, the reduced precipitation forcing in MERRA-Land counteracts the reduced

469   interception loss, typically resulting in somewhat drier or unchanged root zone soil moisture

470   conditions in MERRA-Land (for example in Southeast Asia, in Indonesia, along the tropical

471   South American and African coasts, and in northern Australia).

472   To address the relative realism of the MERRA and MERRA-Land soil moisture estimates, we

473   now validate them against in situ observations taken between 2002 and 2009 in the continental

474   US (Figure 1, section 2b). Our analysis focuses on skill in terms of the anomaly time series

475   correlation coefficient R (Appendix A). Figure 7 shows that for MERRA estimates, the average

476   anomaly skill at pentad time scales is R=0.49 for surface soil moisture (across 98 sites) and

477   R=0.47 for root zone soil moisture (across 85 sites). For MERRA-Land, the anomaly R values

478   increase to R=0.56 for surface and R=0.54 for root zone soil moisture, a net gain of ΔR~0.07

479   over the MERRA R values. Approximate 95% confidence intervals, also shown in Figure 7, are

480   ΔR$\leq\pm$0.01 (Appendix A). The improvements in the MERRA-Land estimates are therefore

481   statistically significant.

482   For comparison, Figure 7 also shows the skill of ERA-I soil moisture estimates (section 2b).

483   ERA-I skill is R=0.58 for surface and R=0.51 for root zone soil moisture. Like MERRA-Land,

484   ERA-I is significantly more skillful than MERRA, but ERA-I does not perform quite as well as

485   MERRA-Land for root zone soil moisture. The ERA-I skill for surface soil moisture is higher

486   than that of MERRA-Land, presumably because the surface layer depth (0-7 cm) of ERA-I better

487   matches the in situ sensing depth (5 cm); MERRA and MERRA-Land use a much shallower (0-2

488   cm) surface layer. Additional analysis (not shown) reveals that most of the improvements in soil

489   moisture skill from MERRA to MERRA-Land can be attributed to the GPCP-based precipitation

490   corrections. The soil moisture skill (in terms of anomaly R) is only weakly sensitive to the

491   changes in the canopy interception parameters of the land model.

492

493   24

494 *c. Runoff*

495 We used naturalized streamflow measurements taken between 1989 and 2009 for 18 basins in the

496 US (Table 1, section 2b) to evaluate runoff estimates. Figure 8 summarizes the skill (anomaly R)

497 at seasonal time scales (Appendix A) for the 9 larger basins and the (area-weighted) average for

498 the 9 smaller basins with areas less than 40,000 km^2 (Table 1). Skill values for MERRA runoff

499 in the larger basins range from R=0.48 for the Arkansas-Red at Arthur City to R=0.83 for the

500 Missouri at Hermann. Because of the 3-month smoothing used here (Appendix A) and because

501 there are typically only 15 years of overlap between the streamflow observations and the

502 reanalysis runoff estimates (Table 1), the 95 % confidence intervals for the R values are large

503 (between ΔR~±0.1 and ΔR~±0.2 for individual basins). MERRA and MERRA-Land, in general,

504 have comparable skill, with three exceptions: MERRA-Land skill is significantly higher than

505 MERRA skill for the Ohio at Metropolis, the Upper Mississippi at Grafton, and the Arkansas-

506 Red at Arthur City.

507 Figure 8 also shows that the skill values for ERA-I are typically lower than those of MERRA and

508 MERRA-Land except for the Ohio at Metropolis, the Upper Mississippi at Grafton, and the

509 Arkansas-Red at Arthur City where ERA-I skill is between that of MERRA and MERRA-Land.

510 ERA-I skill is significantly worse than that of the other estimates for the Milk at Fort Peck Dam

511 and for the average over the 9 small basins. The lower skill of ERA-I is most likely due to the

512 coarser (~1.5 degree) horizontal resolution of the publicly available ERA-I estimates.

513 The revisions to the Catchment model parameters have a small but almost always positive

514 impact. Table 1 shows that in all basins except one small watershed (Yakima near Parker) the R

515   values for the revised Catchment model forced with MERRA surface meteorological data are

516   larger than those of MERRA. While the improvements are not statistically significant, the fact

517   that they occur in so many basins is suggestive of improved hydrological simulation resulting

518   from the improved canopy throughfall rates. However, the significant improvements in

519   MERRA-Land over MERRA noted above are dominated by the positive impact of the GPCP-

520   based precipitation corrections.

521   *d. Snow*

522   We first evaluate the skill of MERRA and MERRA-Land snow depth estimates against in situ

523   measurements taken between 2002 and 2009 at 583 WMO stations in the northern hemisphere

524   (section 2b). The station-average skill (pentad anomaly R; see Appendix A) of snow depth

525   estimates is R=0.56 for MERRA and R=0.59 for MERRA-Land (Table 3). While modest, the

526   skill increase for MERRA-Land is nevertheless statistically significant. An approximate 95%

527   confidence interval for the station-average R value is less than $\Delta R \leq \pm 0.01$ (see Appendix A for

528   details).

529   Errors in modeled snow depth estimates can be caused by errors in the land surface forcing data

530   and by errors in the modeling of snow density. The snow depth bias error is -1.0 cm for

531   MERRA and +5.8 cm for MERRA-Land when averaged over the WMO stations (Table 3).

532   Similarly, station-average snow depth RMSE is 20.1 cm for MERRA and 24.3 cm for MERRA-

533   Land (Table 3). The changes in bias and RMSE (and anomaly R) between MERRA and

534   MERRA-Land are primarily due to the GPCP-based precipitation corrections and are not related

535   to the snow parameter changes (not shown). The snow depth bias may be higher in MERRA-

536   Land because the precipitation gauge undercatch may have been overcorrected in the GPCP

537   precipitation in northern high latitudes (Swenson 2010). A potential bias in the WMO snow

538   depth observations, however, offers another explanation. Most WMO snow depth observations

539   are collected in open areas (such as airports) that are subject to wind-blown snow redistribution.

540   Snow at WMO stations thus tends to be shallower and melt earlier than in surrounding terrain

541   (Brown et al. 2003), which would imply a negative bias in the WMO measurements (relative to

542   the larger-scale conditions).

543   Additional insights can be gained by comparing the MERRA and MERRA-Land snow fields

544   against the CMC snow analysis (section 2b). The CMC product provides a spatially complete

545   estimate of daily northern hemisphere snow depths, conditioned on in situ measurements and

546   aviation reports. Figure 9a maps the skill (pentad anomaly R) of MERRA-Land snow depth

547   versus the CMC product for the period from September 1998 to September 2009. The highest

548   skill values are generally found in southern Siberia and across large portions of Canada and the

549   United States, whereas lower skills are typically found in northern Siberia, the Tibetan plateau,

550   the Canadian Arctic, and in portions of Alaska. For reference, Figure 9c shows the spatial

551   density of in situ snow depth observations that contribute to the CMC snow analysis, based on all

552   stations that were used at least once across the study period. Since only a fraction of these

553   stations are typically used in any given daily analysis, the density map can be thought of as an

554   upper limit.

555   A comparison of Figures 9a and 9c shows that MERRA-Land and CMC snow depth estimates

556   tend to disagree most when the CMC data are based on very few in situ snow depth observations

557   (for example, the high northern latitudes and the Tibetan plateau). That is, the regions of low or

27

558   even negative correlation coincide with areas where actual snow depths are largely unknown.

559   Figure 9c also resembles the density of precipitation gauges used for conditioning the GPCP

560   estimates and that of the radiosonde observations available for assimilation into MERRA (not

561   shown). This implies that MERRA-Land (and MERRA) estimates are based on fewer

562   conventional observations and are thus likely less accurate wherever CMC snow depths are less

563   accurate.

564   The geographic skill pattern for MERRA snow depths (not shown) is similar to that of MERRA-

565   Land estimates (Figure 9a). Similar geographic patterns are also evident in the skill analysis

566   against the WMO in situ snow depth measurements (not shown), which is not surprising because

567   the CMC product is conditioned on WMO snow depth measurements when and where available.

568   Area-weighted pentad anomaly skill versus CMC snow depth is R=0.51 for MERRA and R=0.50

569   for MERRA-Land (Table 3). If the skill average is taken only over CMC grid cells that contain

570   the 583 WMO stations used above, snow depth skill increases to R=0.60 for MERRA and

571   R=0.61 MERRA-Land, which is consistent with the skill values assessed directly against the

572   WMO measurements (Table 3).

573   The ERA-I snow depth analysis is largely based on the same in situ snow depth observations

574   used for conditioning the CMC product, although the analysis update is different between the

575   two products (section 2b).   Given that these in situ observations were not assimilated into

576   MERRA, it is no surprise that ERA-I anomaly snow depth correlations versus CMC (Figure 9b)

577   are higher than those of MERRA-Land (or MERRA) versus CMC in eastern Europe, the western

578   half of Russia, and the eastern US, that is, in regions with a dense network of in situ snow depth

579   stations (Figure 9c). Across the 583 WMO stations, the average skill (pentad anomaly R) of

28

580   ERA-I snow depth is R=0.63, which is slightly higher than MERRA-Land and significantly

581   higher than MERRA skill (see above). However, across the northern hemisphere the average

582   correlation of ERA-I snow depth versus CMC is only R=0.39 (Table 3) and thus considerably

583   lower than that of MERRA-Land (or MERRA) versus CMC. Lower correlations can be seen in

584   eastern Siberia, northern Canada, and Alaska (Figure 9b). Because there are few stations in these

585   regions, it is not possible to tell which of the products is closer to reality.

586   By combining CMC snow depths with state-of-the-art snow density estimates (Sturm et al. 2010;

587   section 2b) we extended our evaluation to SWE, a quantity of more relevance to hydrology. The

588   area-weighted skill of SWE pentad anomalies is R=0.49 for MERRA, R=0.49 for MERRA-

589   Land, and R=0.38 for ERA-I (Table 3), comparable to the anomaly R values for snow depth.

590   The spatial pattern of the SWE skills (not shown) is very similar to that of snow depth skills

591   (Figure 9a,b). Table 3 also lists the bias and RMSE values for MERRA, MERRA-Land, and

592   ERA-I snow depth and SWE versus CMC estimates. By and large, these values are consistent

593   with the snow depth bias and RMSE values versus WMO.

594

595

5.   Discussion and conclusions

Reanalysis estimates of surface meteorological forcings and land surface fields such as snow and soil moisture have proven useful for research into the global water and energy cycles, seasonal climate forecasting, and hydrology. In this paper we assess the skill of soil moisture, snow, and runoff estimates from MERRA against a variety of in situ observations. We also introduce an improved land surface data set, MERRA-Land, motivated by limitations in MERRA surface meteorological fields, specifically errors in the long-term climatology, the diurnal cycle, and the intensity of precipitation (Figures 2 and 3). Such deficiencies are indeed typical of global reanalyses and adversely affect the simulation of land surface hydrology. MERRA-Land is a "replay" of the MERRA system's land surface component that benefits from corrections to the precipitation forcing at the pentad scale (using the GPCP v2.1 pentad product) and from revisions to the Catchment model's interception parameters designed to counterbalance known precipitation deficiencies at the sub-diurnal scale. The MERRA-Land data products will be made available to the community.

We focus our skill analysis on time series correlation coefficients (versus observations) of pentad average anomalies (soil moisture, snow) or 3-month average anomalies (runoff). Note that because we examine anomalies here, we avoid extracting 'trivial' skill from the simulation of the mean seasonal cycle. Generally, the skill of MERRA and MERRA-Land estimates of soil moisture and runoff is comparable to that of ERA-I estimates. Moreover, snow depth and SWE compare well against in situ observations and the state-of-the-art CMC snow analysis. Average (anomaly) skill levels for MERRA and MERRA-Land surface hydrological variables generally range from R~0.5 to R~0.9 (Figures 7, 8, and 9). The skill of MERRA-Land estimates is higher

618   than that of MERRA estimates by ΔR~0.07 for soil moisture (Figure 7) and ΔR~0.03 for snow

619   depth (Table 3), differences that are statistically significant at the 5% level. Moreover, MERRA-

620   Land runoff skill is significantly better than that of MERRA for three of the nine large basins

621   examined here (Table 1, Figure 8). The skill improvements for these variables are typically

622   derived from the GPCP-based precipitation corrections; the revisions to the Catchment model

623   parameters contribute a smaller fraction to the overall improvement. The revised interception

624   model parameters, however, considerably improve the average interception loss fraction (Figure

625   4) and contribute to more realistic latent heat fluxes (Figure 5) in MERRA-Land.

626   Future reanalysis efforts should include the assimilation of land surface observations. For

627   example, Liu et al. (2011) find that the assimilation of surface soil moisture retrievals provides

628   important information that is largely independent of that provided by the precipitation

629   observations. Soil moisture data assimilation has in fact matured to the point where few

630   technical obstacles remain for a long-term soil moisture analysis, though we note that X- or C-

631   band passive or active microwave observations are not available for the entire satellite era (1979-

632   present). The assimilation of screen-level air temperature and humidity observations has been

633   operational at a number weather centers and is used in ERA-I (section 2b). For the assimilation

634   of satellite-based land surface temperature data, abundant observations are available throughout

635   the satellite era, though appropriate assimilation approaches are considerably less mature

636   (Reichle et al. 2010). The assimilation of snow cover fraction (Zaitchik and Rodell 2009) shows

637   promise, and while MERRA does not contain a snow analysis, most weather centers have been

638   assimilating satellite snow cover observations and in situ snow depth measurements for many

639   years (for example, Drusch et al. 2004). Even though current-generation satellite retrievals of

640   SWE do not appear to be accurate enough for use in land assimilation, emerging dynamic

641   retrieval approaches may provide useful information (Tedesco et al. 2010), and progress has

642   been made towards a radiance-based SWE analysis (Durand and Margulis 2008). Total water

643   storage information from GRACE has been successfully assimilated into a land surface model

644   (Zaitchik et al. 2008). Advances in the utilization of all of these land data sources are

645   continually proceeding. It seems reasonable to predict that next-generation estimates of global

646   land surface hydrological fields will indeed be based on a comprehensive land surface analysis.

647

648

667

Appendix A: Skill metric

Bias is a common problem when validating land model estimates representing scales of ~10-50 km against point-scale in situ measurements such as the soil moisture and snow depth observations used here; see for instance (Reichle et al. 2004). For soil moisture, the discrepancy between the modeled layer depths and the depths at which in situ sensors are installed can lead to additional bias errors. Specifically, Catchment model surface soil moisture covers the top 2 cm of the soil column while the in situ surface soil moisture observations were taken at 5 cm depth. Moreover, Catchment model root zone soil moisture covers the top 1 m of the soil profile and is validated with a depth-weighted average of the SCAN sensors at 5 cm, 10 cm, and 20 cm, because quality-controlled SCAN data at 50 cm and 100 cm were too sparse and intermittent (Reichle et al. 2007; Liu et al. 2011).

Fortunately, temporal variations (in a percentile sense) are typically more important for model-based applications (Entekhabi et al. 2010). We therefore first compute the climatological seasonal cycle over the period of interest (separately for each data product), obtain anomalies by subtracting this climatology from the time series, and finally assess skill in terms of correlation coefficients (R values). For soil moisture and snow depth we constructed pentad average anomaly time series (because GPCP precipitation estimates are pentad averages). For runoff, we constructed smoothed anomalies by applying a 3-month moving average to the anomalies (because MERRA and ERA-I lack routing schemes). For the soil moisture skill analysis we excluded from the computation of the R values the times and locations for which the soil was frozen. Similarly, for the snow skill analysis we excluded times and locations for which WMO (or CMC) indicated snow free conditions.

690 For the results presented here we first computed anomaly R values for each site (or grid cell) and

691 then computed the average skill by averaging the R values across all sites. Common masks and

692 minimum data requirements were applied to ensure consistent and sensible estimates of the

693 climatological seasonal cycle on which the anomalies are based. For soil moisture and snow, we

694 also required a minimum of 50 pentad average anomalies across the multi-year experiment

695 period for computing the anomaly R value.

696 We also computed approximate 95% confidence intervals for the R estimates at each site based

697 on the Fisher Z transform. These confidence intervals depend on the estimated R value and on

698 the number of degrees of freedom, which is approximated here by the number of pentad averages

699 that go into the R computation (for soil moisture and snow). Because of the 3-month smoothing

700 we only assume four degrees of freedom per data year in the runoff skill analysis. The

701 approximate 95% confidence intervals for the average skill estimates across all sites were then

702 computed by averaging the 95% confidence intervals of the N contributing sites and

703 subsequently dividing by the square root of N. It is important to stress that the 95 % confidence

704 intervals computed here are only approximations and may underestimate the true widths of the

705 confidence intervals because temporal error correlations may reduce the number of degrees of

706 freedom below the numbers assumed here.

707

708

709   **References**

710   Adler, R. F., and Coauthors, 2003: The Version-2 Global Precipitation Climatology Project

711   (GPCP) Monthly Precipitation Analysis (1979–Present). *J. Hydrometeor*, **4**, 1147–1167.

712   Berg, A. A., J. S. Famiglietti, M. Rodell, R. H. Reichle, U. Jambor, S. L. Holl, and P. R. Houser,

713   2005: Development of a Hydrometeorological Forcing Data Set for Global Soil Moisture

714   Estimation. *International Journal of Climatology,* **25**, 1697-1714, doi:10.1002/joc.1203.

715   Bosilovich, M. G., F. R. Robertson, and J. Chen, 2011: Global Energy and Water Budgets in

716   MERRA. *J. Climate*, submitted.

717   Brasnett, B., 1999: A Global Analysis of Snow Depth for Numerical Weather Prediction. *J. Appl.*

718   *Meteorol.*, **38**, 726-740.

719   Brown, R. D. and B. Brasnett, 2010: *Canadian Meteorological Centre (CMC) Daily Snow Depth*

720   *Analysis Data.* Environment Canada, 2010. Boulder, Colorado USA. National Snow and Ice

721   Data Center, Digital Media.

722   Brown, R. D., B. Brasnett, and D. Robinson, 2003: Gridded North American monthly snow

723   depth and snow water equivalent for GCM evaluation. *Atmos. Ocean*, **41**, 1-14.

724   Brown, R., C. Derksen, and L. Wang, 2010: A multi-data set analysis of variability and change

725   in Arctic spring snow cover extent, 1967–2008. *J. Geophys. Res.*, **115**, D16111,

726   doi:10.1029/2010JD013975.

727   Burke, E. J., R. H. J. Perry, and S. J. Brown, 2010: An extreme value analysis of UK drought and

728   projections of change in the future. *J. Hydrology*, **388,** 131-143.

36

729   Dai, A., 2006: Precipitation Characteristics in Eighteen Coupled Climate Models. *J. Climate*, **19**,

730   4605–4630.

731   Dai, A., and K. E. Trenberth, 2002: Estimates of Freshwater Discharge from Continents:

732   Latitudinal and Seasonal Variations. *J. Hydrometeor*, **3**, 660-687.

733   Decker, M., M. Brunke, Z. Wang, K. Sakaguchi, X. Zeng, and M. G. Bosilovich, 2011:

734   Evaluation of the Reanalysis Products from GSFC, NCEP, and ECMWF Using Flux Tower

735   Observations. *J. Climate*, submitted.

736   Dee, D. P., E. Källén, A. J. Simmons, and L. Haimberger, 2010: Comments on "Reanalyses

737   suitable for characterizing long-term trends". *Bull. Amer. Meteor. Soc.*, **92**, 65-70,

738   10.1175/2010BAMS3070.1.

739   Dee, D. P. and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of

740   the data assimilation system. *Quart. J. Roy. Meteor. Soc.,* submitted.

741   Dirmeyer, P. A., and L. Tan, 2001: A multi-decadal global land-surface data set of state variables

742   and fluxes. COLA Technical Report 102 [Available from the Center for Ocean-Land-

743   Atmosphere Studies, 4041 Powder Mill Road, Suite 302, Calverton, MD 20705 USA], 43 pp.

744   Dirmeyer, P. A., X. Gao, M. Zhao, Z. Guo, T. Oki and N. Hanasaki, 2006: The Second Global

745   Soil Wetness Project (GSWP-2): Multi-model analysis and implications for our perception of the

746   land surface. *Bull. Amer. Meteor. Soc.*, **87**, 1381-1397.

747   Ducharne, A., R. D. Koster, M. J. Suarez, M. Stieglitz, and P. Kumar, 2000: A catchment-based

748   approach to modeling land surface processes in a general circulation model 2. Parameter

749 estimation and model demonstration. *J. Geophys. Res.*, **105**, 24,823–24,838,

750 doi:10.1029/2000JD900328.

751 Durand, M., and S. A. Margulis, 2008: Effects of uncertainty magnitude and accuracy on

752 assimilation of multiscale measurements for snowpack characterization. *J. Geophys. Res.*, **113**,

753 D02105, doi:10.1029/2007JD008662.

754 Drusch, M., and P. Viterbo, 2007: Assimilation of Screen-Level Variables in ECMWF's

755 Integrated Forecasting System: A Study on the Impact on the Forecast Quality and Analyzed Soil

756 Moisture. *Monthly Weather Review,* **135,** 300-314.

757 Drusch, M., D. Vasiljevic, and P. Viterbo, 2004: ECMWF's Global Snow Analysis: Assessment

758 and Revision Based on Satellite Observations. *Journal of Applied Meteorology*, **43**, 1282-1294.

759 Entekhabi, D., R. H. Reichle, R. D. Koster, and W. T. Crow, 2010: Performance Metrics for Soil

760 Moisture Retrievals and Application Requirements. *J. Hydrometeor.*, **11**, 832-840,

761 doi:10.1175/2010JHM1223.1.

762 Guo, Z., P. A. Dirmeyer, Z. Hu, X. Gao, M. Zhao, 2006: Evaluation of the Second Global Soil

763 Wetness Project soil moisture simulations: 2. Sensitivity to external meteorological forcing. *J.*

764 *Geophy. Res.*, **111**, D22S03, doi:10.1029/2006JD007845.

765 Huffman, G. J., R. F. Adler, D. T. Bolvin, and G. Gu 2009: Improving the global precipitation

766 record: GPCP Version 2.1. *Geophys. Res. Lett.*, **36**, L17808, doi:10.1029/2009GL040000.

767 Jimenez, C., C. Prigent, and Coauthors, 2011: Global inter-comparison of 12 land surface heat

768 flux estimates. *J. Geophys. Res.*, **116**, D02102, doi:10.1029/2010JD014545.

38

769   Kalnay, E. and 21 co-authors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer.*

770   *Meteor. Soc.*, **77**, 437-471.

771   Kanamitsu, M., W. Ebisuzaki, J. Woollen, S.-K. Yang, J. J. Hnilo, M. Fiorino, G. L. Potter,

772   2002: NCEP–DOE AMIP-II Reanalysis (R-2). *Bull. Amer. Meteor. Soc.,* **83**, 1631-1643.

773   Koster, R. D., M. J. Suarez, A. Ducharne, M. Stieglitz, and P. Kumar, 2000: A catchment-based

774   approach to modeling land surface processes in a general circulation model 1. Model structure.

775   *J. Geophys. Res.*, **105**(D20), 24,809-24,822.

776   Koster, R. D., S. P. P. Mahanama, B. Livneh, D. Lettenmaier, and R. H. Reichle, 2010: Skill in

777   Streamflow Forecasts Derived from Large-Scale Estimates of Soil Moisture and Snow. *Nature*

778   *Geoscience,* **3**, 613-616, doi:10.1038/ngeo944.

779   Liu, Q., and Coauthors, 2011: The contributions of precipitation and soil moisture observations

780   to the skill of soil moisture estimates in a land data assimilation system. *J. Hydrometeor*.,

781   doi:10.1175/JHM-D-10-05000, in press.

782   Mahanama, S. P. P., B. Livneh, R. D. Koster, D. Lettenmaier, and R. H. Reichle, 2011: Soil

783   Moisture, Snow, and Seasonal Streamflow Forecasts in the United States. *J. Hydrometeor.*,

784   submitted.

785   Miralles, D. G., J. H. Gash, T. R. H. Holmes, R. A. M. de Jeu, and A. J. Dolman, 2010: Global

786   canopy interception from satellite observations. *J. Geophys. Res.*, **115**, D16122,

787   doi:10.1029/2009JD013530.

788   Onogi, K., and Coauthors, 2007: The JRA-25 reanalysis. *J. Meteor. Soc. Japan*, **85**, 369-432.

789   Reichle, R. H., R. D. Koster, J. Dong, and A. A. Berg, 2004: Global Soil Moisture from Satellite

790   Observations, Land Surface Models, and Ground Data: Implications for Data Assimilation. *J.*

791   *Hydrometeor.*, **5**, 430-442.

792   Qian, T., A. Dai, K. E. Trenberth, K. W. Oleson, 2006: Simulation of Global Land Surface

793   Conditions from 1948 to 2004. Part I: Forcing Data and Evaluations. *J. Hydrometeor*, **7**, 953–

794   975.

795   Reichle, R. H., R. D. Koster, P. Liu, S. P. P. Mahanama, E. G. Njoku, and M. Owe, 2007:

796   Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave

797   Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel

798   Microwave Radiometer (SMMR). *J. of Geophys. Res.*, **112**, D09108,

799   doi:10.1029/2006JD008033.

800   Reichle, R. H., S. V. Kumar, S. P. P. Mahanama, R. D. Koster, and Q. Liu, 2010: Assimilation of

801   satellite-derived skin temperature observations into land surface models. *J. Hydrometeor.*, **11**,

802   1103-1122, doi:10.1175/2010JHM1262.1.

803   Rienecker, M. M., and Coauthors, 2011: An overview of MERRA. *J. Climate*,

804   doi:10.1175/JCLI-D-11-00015, in press.

805   Saha, S., and Coauthors, 2006: The NCEP Climate Forecast System. *J. Climate*, **19**, 3483–3517.

806   Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer.*

807   *Meteor. Soc.*, **91**, 1015-1057, doi: 10.1175/2010BAMS3001.1.

808   Sakaguchi, K., and X. Zeng, 2009: Effects of soil wetness, plant litter, and under-canopy

809   atmospheric stability on ground evaporation in the Community Land Model (CLM3.5). *J.*

810   *Geophys. Res.*, **114**, D01107, doi:10.1029/2008JD010834.

811   Schaefer, G.L., M.H. Cosh and T.J. Jackson, 2007: The USDA Natural Resources Conservation

812   Service Soil Climate Analysis Network (SCAN). *J. Atmos. Oceanic Technol.*, **24**, 2073-2077.

813   Szczypta, C., and Coauthors, 2011: Verification of the new ECMWF ERA-Interim reanalysis

814   over France. *Hydrol. Earth Syst. Sci.*, **15**, 647-666, doi:10.5194/hess-15-647-2011.

815   Sheffield, J., Goteti, G., Wood, E. F., 2006: Development of a 50-year high-resolution global

816   dataset of meteorological forcings for land surface modeling. *J. Climate*, **19**, 3088-3111.

817   Sheffield, J., and E. F. Wood, 2008: Global Trends and Variability in Soil Moisture and Drought

818   Characteristics, 1950–2000, from Observation-Driven Simulations of the Terrestrial Hydrologic

819   Cycle. *J. Climate*, **21**, 432–458.

820   Stieglitz, M., A. Ducharne, R. Koster, and M. Suarez, 2001: The Impact of Detailed Snow

821   Physics on the Simulation of Snow Cover and Subsurface Thermodynamics at Continental

822   Scales. *J. Hydrometeor.*, **2**, 228-242.

823   Sturm, M., J. Holmgren, and G. E. Liston, 1995: A Seasonal Snow Cover Classification System

824   for Local to Global Applications. *J. Climate*, **8**, 1261-1283.

825   Sturm, M., B. Taras, G. E. Liston, C. Derksen, T. Jonas, and J. Lea, 2010: Estimating Snow

826   Water Equivalent Using Snow Depth Data and Climate Classes, *J. Hydrometeor.*, **11**, 1380–

827   1394, doi: 10.1175/2010JHM1202.1.

828   Su, F., and D. P. Lettenmaier, 2009: Estimation of the Surface Water Budget of the La Plata

829   Basin. *J. Hydrometeor*, **10**, 981–998.

830   Su, H., Z.-L. Yang, R. E. Dickinson, C. R. Wilson, and G.-Y. Niu, 2010: Multisensor snow data

831   assimilation at the continental scale: The value of Gravity Recovery and Climate Experiment

832   terrestrial water storage information. *J. of Geophys. Res*., **115**, D10104,

833   doi:10.1029/2009JD013035.

834   Swenson, S., 2010: Assessing High-Latitude Precipitation from Global Precipitation Analyses

835   Using GRACE. *J. Hydrometeor.*, **11**, 405-420, doi: 10.1175/2009JHM1194.1.

836   Tedesco, M. and J. Miller, 2010: Co-Registered AMSR-E, QuikSCAT, and WMO Data.

837   Boulder, Colorado USA: National Snow and Ice Data Center. Digital media.

838   Tedesco, M., R. H. Reichle, A. Loew, T. Markus, and J. L. Foster, 2010: Dynamic Approaches

839   for Snow Depth Retrieval From Spaceborne Microwave Brightness Temperature. *IEEE*

840   *Transactions on Geosciences and Remote Sensing*, **48**, 1955-1967,

841   doi:10.1109/TGRS.2009.2036910.

842   Thorne, P. W., and R. S. Vose, 2010: Reanalyses suitable for characterizing long-term trends.

843   *Bull. Amer. Meteor. Soc.*, **91**, 353-361, doi:10.1175/2009BAMS2858.1.

844   Uppala, S.M., and Coauthors, 2005: The ERA-40 re-analysis. *Quart. J. R. Meteorol. Soc.*, **131**,

845   2961-3012.doi:10.1256/qj.04.176.

846   Xie P. and P. A. Arkin, 1997: Global Precipitation: A 17-year monthly analysis based on gauge

847   observations, satellite estimates, and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78**,

848   2539-2558.

849   Xie P., J. E. Janowiak, P. A. Arkin, R. Adler, A. Gruber, R. Ferraro, G. J. Huffman and S. Curtis,

850   2003: GPCP pentad precipitation analyses: an experimental dataset based on gauge observations

851   and satellite estimates. *J. Climate*, **16**, 2197-2214.

852   Yi, Y., J. Kimball, L. Jones, R. H. Reichle, and K. McDonald, 2011: Evaluation of MERRA land

853   surface estimates in preparation for the Soil Moisture Active Passive mission. *J. Climate*,

854   doi:10.1175/2011JCLI4034.1, in press.

855   Zaitchik, B. F., and M. Rodell, 2009: Forward-looking assimilation of MODIS-derived snow-

856   covered area into a land surface model. *J. Hydrometeor.*, **10**, 130–148.

857   Zaitchik, B. F., M. Rodell, and R. H. Reichle, 2008: Assimilation of GRACE terrestrial water

858   storage data into a land surface model: Results for the Mississippi River basin. *J. Hydrometeor.*,

859   **9**, 535-548, doi:10.1175/2007JHM951.1.

860   Zhao, M., S. W. Running, and R. R. Nemani, 2006: Sensitivity of Moderate Resolution Imaging

861   Spectroradiometer (MODIS) terrestrial primary production to the accuracy of meteorological

862   reanalyses. *J. Geophys. Res.*, 111, G01002, doi:10.1029/2004JG000004.

863

864   Table 1. Characteristics of the basins examined in this study and skill (anomaly R) of 3-month smoothed streamflow. "Revised

865   CLSM" skills are for an integration of the Catchment model with surface meteorological forcing from MERRA and Catchment model

866   parameters from MERRA-Land (Table 2).

867

Basin	River Name	Station Name	Basin Area	Latitude	Longitude	Period	Anomaly R Value			
			km²	degree N	degree W		MERRA	Revised CLSM	MERRA-Land	ERA-I
1	Missouri	Hermann (incl. basins 2, 6, 9 and 10d)	1,357,667	38.71	92.75	1989-1997	0.83	0.86	0.85	0.79
2	Missouri	Ft Randall Dam (incl. basins 6, 9 and 10d)	691,530	43.07	98.55	1989-2009	0.74	0.75	0.66	0.59
3	Ohio	Metropolis	525,760	37.15	88.74	1989-2010	0.71	0.72	0.93	0.77
4	Upper Mississippi	Grafton	443,660	38.90	90.30	1989-2010	0.64	0.65	0.89	0.81
5	Colorado	Lees Ferry (incl. basins 10a and 10e)	278,070	36.87	111.58	1989-2003	0.53	0.60	0.52	0.46
6	Milk	Fort Peck Dam (incl. basin 10d)	149,070	48.04	106.36	1989-2009	0.79	0.82	0.77	0.51
7	Arkansas	Ralston	121,341	36.50	98.73	1989-2008	0.65	0.66	0.70	0.54
8	Arkansas-Red	Arthur City	99,961	33.88	95.50	1989-2001	0.48	0.55	0.89	0.63
9	Missouri	Garrison Reservoir	89,355	47.39	101.39	1989-2003	0.56	0.58	0.58	0.54
10a	Green	Greendale	39,452	40.91	109.42	1989-2003	0.61	0.61	0.57	0.42
10b	Potomac	Point of Rocks	25,000	39.27	77.54	1989-1996	0.87	0.88	0.95	0.83
10c	Sacramento	Bend Bridge	23,051	40.29	122.19	1989-2003	0.93	0.94	0.93	0.91
10d	Musselshel	Moseby	20,321	46.99	107.89	1989-2003	0.69	0.73	0.65	0.32
10e	Gunnison	near Grand Junction	19,958	38.98	108.45	1989-2003	0.53	0.60	0.51	0.25
10f	Rio Puerco	Bernardo	19,036	34.41	106.85	1989-2003	0.20	0.25	0.45	0.26
10g	Yakima	near Parker	9,479	46.50	120.44	1989-2003	0.63	0.60	0.69	0.59
10h	Tuolumne	La Grange Dam	4,337	37.67	120.44	1989-2003	0.66	0.67	0.67	0.70
10i	San Joaquin	Mokelunme Hill	1,863	38.31	120.72	1989-2003	0.69	0.70	0.68	0.72
Area-weighted average over small basins (10a-i)			*n/a*	*n/a*	*n/a*	*n/a*	*0.65*	*0.67*	*0.68*	*0.52*

868
869

870 Table 2. Catchment land surface model parameter changes between MERRA and the revised

871 Catchment model used in MERRA-Land. SATCAP is computed as a fraction of leaf area index

872 (LAI).

873

Parameter	Description	Units	MERRA	MERRA-Land
SATCAP	Capacity of canopy interception reservoir	kg/m^2	1.0*LAI	0.2*LAI
FWETL	Areal fraction of canopy leaves onto which large-scale precipitation falls	dimensionless	1.0	0.02
FWETC	Areal fraction of canopy leaves onto which convective precipitation falls	dimensionless	0.2	0.02
WEMIN	Minimum SWE in snow-covered area fraction	kg/m^2	13	26
DZ1MAX	Maximum depth of uppermost snow layer	m	0.05	0.08

874

875

876   Table 3. Skill summary for snow estimates. Anomaly R values vs. WMO measurements at 583

877   stations are provided with approximate 95% confidence intervals.   Skill vs. CMC is area-

878   weighted average over northern hemisphere grid cells (Figure 9a,b).

879

Metric	Units	Dataset	Snow depth		SWE
			vs. WMO	vs. CMC	vs. CMC + Sturm et al. (2010)
Anomaly R	dimension-less	MERRA	0.56±0.01	0.51	0.49
		MERRA-Land	0.59±0.01	0.50	0.49
		ERA-I	0.60±0.01	0.39	0.38
Bias	cm	MERRA	-1.0	-2.3	-1.2
		MERRA-Land	5.8	-0.2	-0.6
		ERA-I	5.2	1.7	0.2
RMSE	cm	MERRA	20.1	9.5	3.7
		MERRA-Land	24.3	12.0	4.4
		ERA-I	25.7	15.0	5.5

880

Figure captions:

882 Figure 1. Locations of SCAN soil moisture measurement sites that were (crosses) used for

883 surface and root zone soil moisture validation (85 sites), (circles) used only for surface soil

884 moisture validation (13 sites), and (dots) not used.

885

886 Figure 2. Annual precipitation (mm d^{-1}) averaged over the period 1981-2008 for (a) MERRA

887 and (b) GPCP. Precipitation differences (MERRA minus GPCP in mm d^{-1}) averaged over (c)

888 1981-2008 and (d) August 1994.

889

890 Figure 3. (Gray lines) Downward shortwave radiation and (black bars) precipitation from (top)

891 MERRA for a grid cell near Gainesville, Florida (centered at 30ºN, 82ºW) and (bottom) in situ

892 observations taken at the US-SP3 FLUXNET site (29.75ºN, 82.16ºW) located within the grid

893 cell.

894

895 Figure 4. 2003-2007 average interception loss fraction (dimensionless) from (a) MERRA, (b)

896 revised Catchment model with MERRA forcing, (c) MERRA-Land, and (d) observations-based

897 estimates from Miralles et al. (2010). Note the different colorbar in (a).

898

899 Figure 5. Average latent heat flux (W m^{-2}) for August 1994 from (a) MERRA, (b) MERRA-

900 Land, and (c) the Jimenez et al. (2011) multi-product average.

901   Figure 6.   Annual average root zone soil moisture (m^3m^{-3}) differences (1981-2008): (a) MERRA

902   minus revised Catchment model forced with MERRA surface meteorology, and (b) MERRA

903   minus MERRA-Land.

904

905   Figure 7.   Skill (pentad anomaly R; dimensionless) of MERRA, MERRA-Land, and ERA-I

906   estimates (2002-2009) versus SCAN in situ surface and root zone soil moisture measurements.

907   Error bars indicate approximate 95% confidence intervals.

908

909   Figure 8.   Seasonal anomaly time series correlation coefficients (dimensionless) for runoff

910   estimates from MERRA, MERRA-Land, and ERA-I for the basins and time periods listed in

911   Table 1.

912

913   Figure 9.   Skill (pentad anomaly R) of (a) MERRA-Land and (b) ERA-I snow depth versus CMC

914   estimates (September 1998 – September 2009). R values that are not statistically different from

915   zero at the 5% significance level are shown in gray. (c) Maximum density of in situ snow depth

916   measurements available for CMC snow analysis.

917

918

919

920   Figure 1. Locations of SCAN soil moisture measurement sites that were (crosses) used for

921   surface and root zone soil moisture validation (85 sites), (circles) used only for surface soil

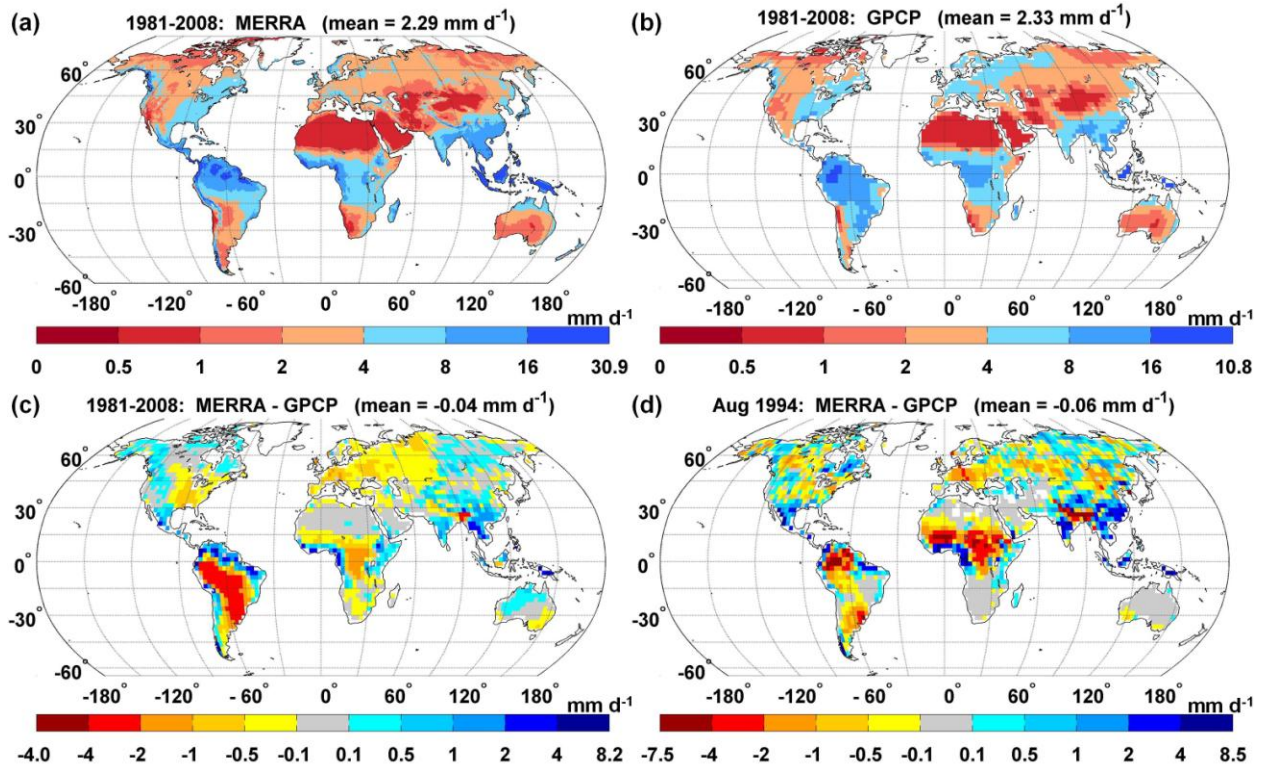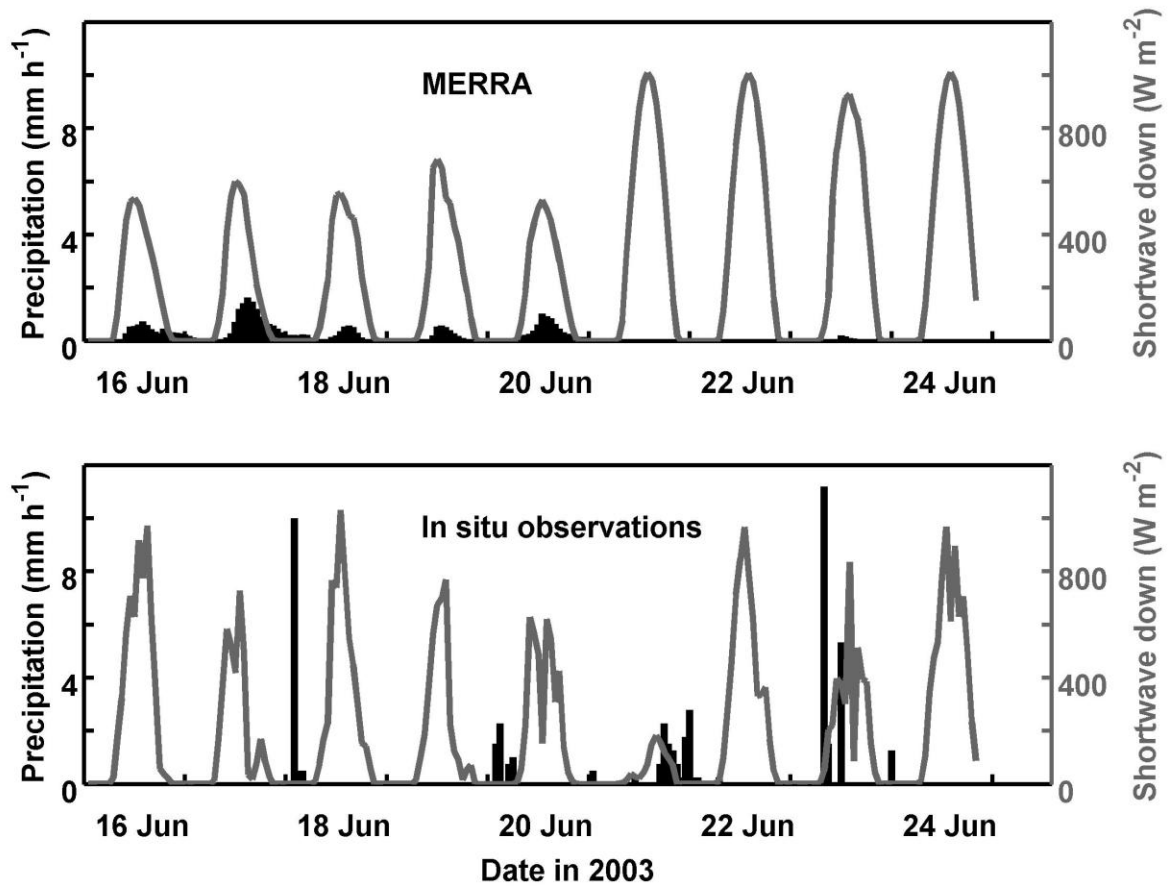922   moisture validation (13 sites), and (dots) not used.

923

924

925

Figure 2. Annual precipitation (mm d^{-1}) averaged over the period 1981-2008 for (a) MERRA

and (b) GPCP.   Precipitation differences (MERRA minus GPCP in mm d^{-1}) averaged over (c)

1981-2008 and (d) August 1994.

929

930
931

Figure 3. (Gray lines) Downward shortwave radiation and (black bars) precipitation from (top)

MERRA for a grid cell near Gainesville, Florida (centered at 30ºN, 82ºW) and (bottom) in situ

observations taken at the US-SP3 FLUXNET site (29.75°N, 82.16°W) located within the grid

cell.

936
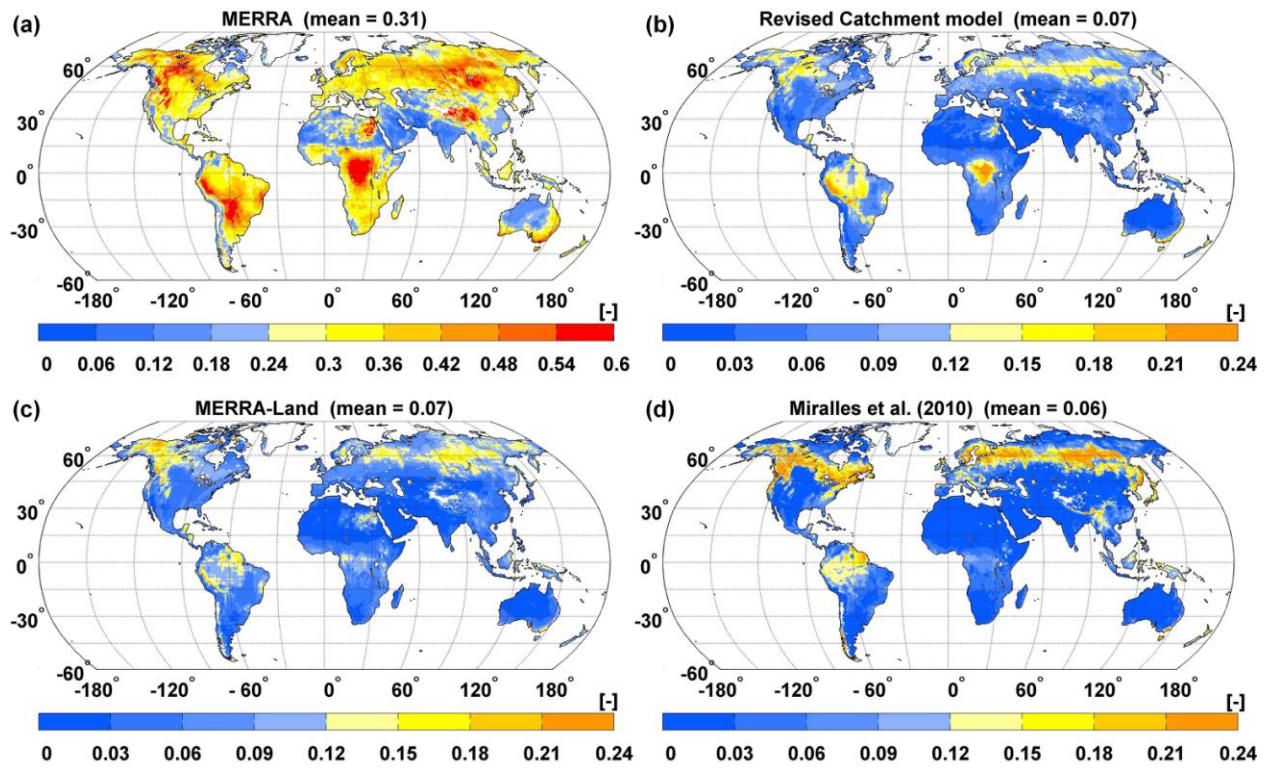
Figure 4. 2003-2007 average interception loss fraction (dimensionless) from (a) MERRA, (b) revised Catchment model with MERRA forcing, (c) MERRA-Land, and (d) observations-based estimates from Miralles et al. (2010). Note the different colorbar in (a).

943

944
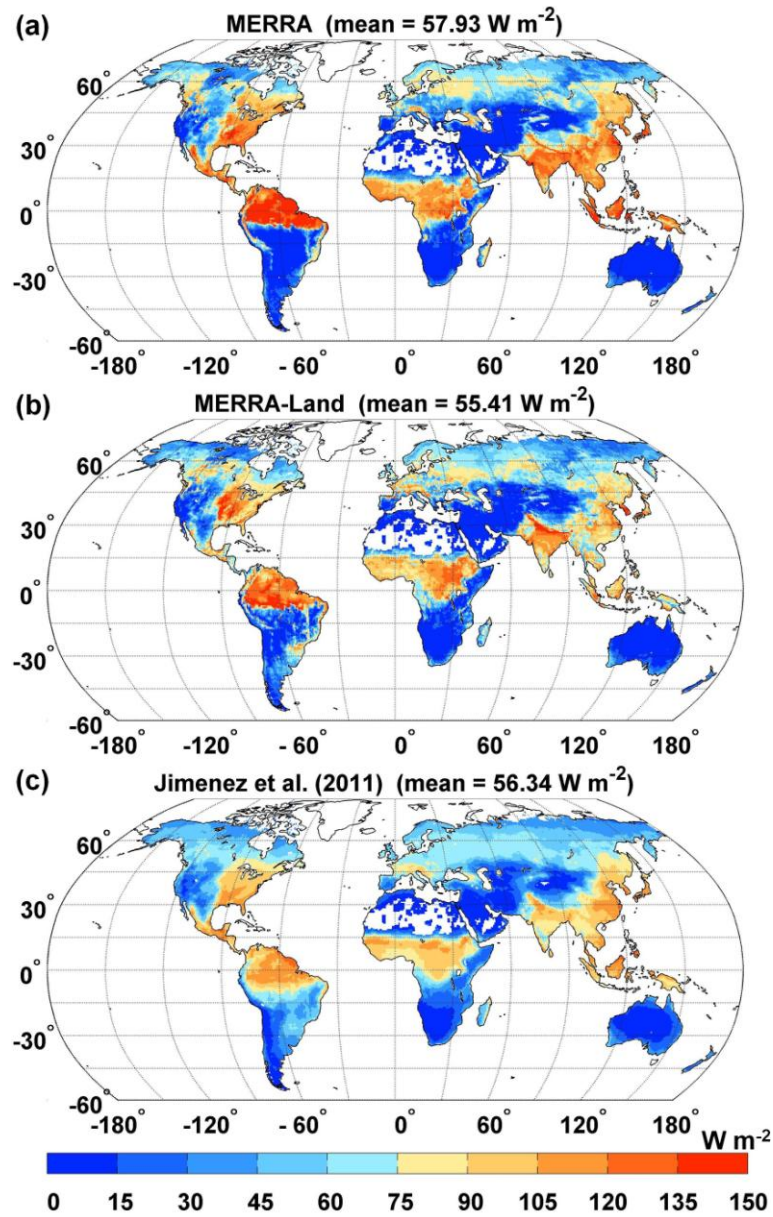
Figure 5. Average latent heat flux (W m^{-2}) for August 1994 from (a) MERRA, (b) MERRA-

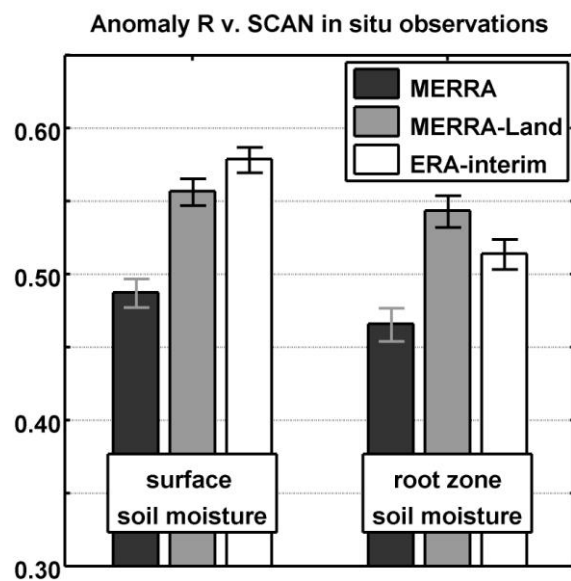Land, and (c) the Jimenez et al. (2011) multi-product average.

947
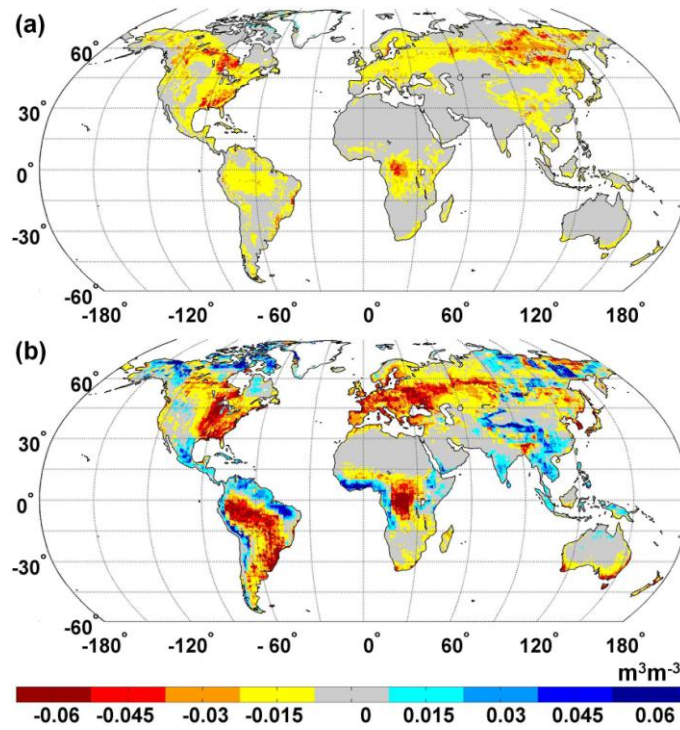
948

949

950   Figure 6. Annual average root zone soil moisture ($m^3 m^{-3}$) differences (1981-2008): (a) MERRA

951   minus revised Catchment model forced with MERRA surface meteorology, and (b) MERRA

952   minus MERRA-Land.

953

Anomaly R v. SCAN in situ observations

954

955

956   Figure 7. Skill (pentad anomaly R; dimensionless) of MERRA, MERRA-Land, and ERA-I

957   estimates (2002-2009) versus SCAN in situ surface and root zone soil moisture measurements.

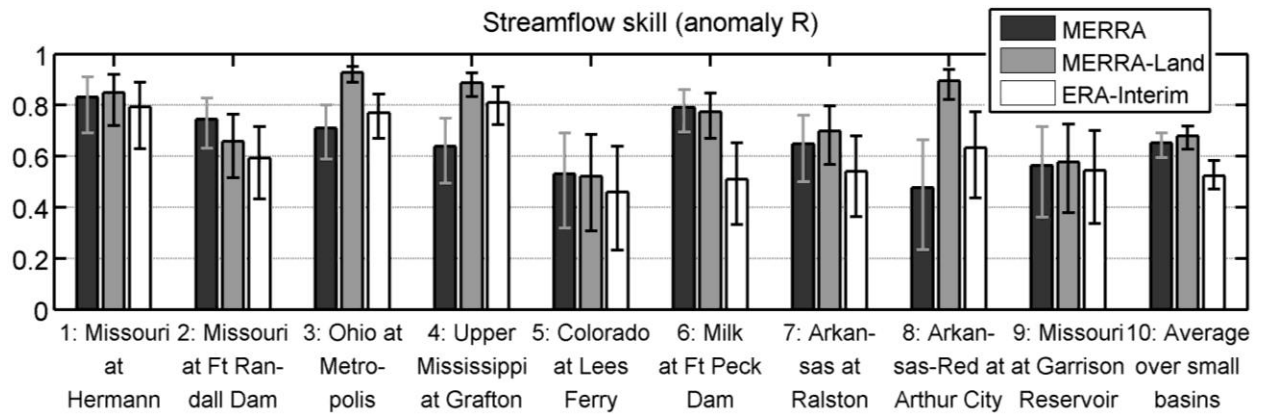958   Error bars indicate approximate 95% confidence intervals.

959

Figure 8. Seasonal anomaly time series correlation coefficients (dimensionless) for runoff estimates from MERRA, MERRA-Land, and ERA-I for the basins and time periods listed in Table 1.
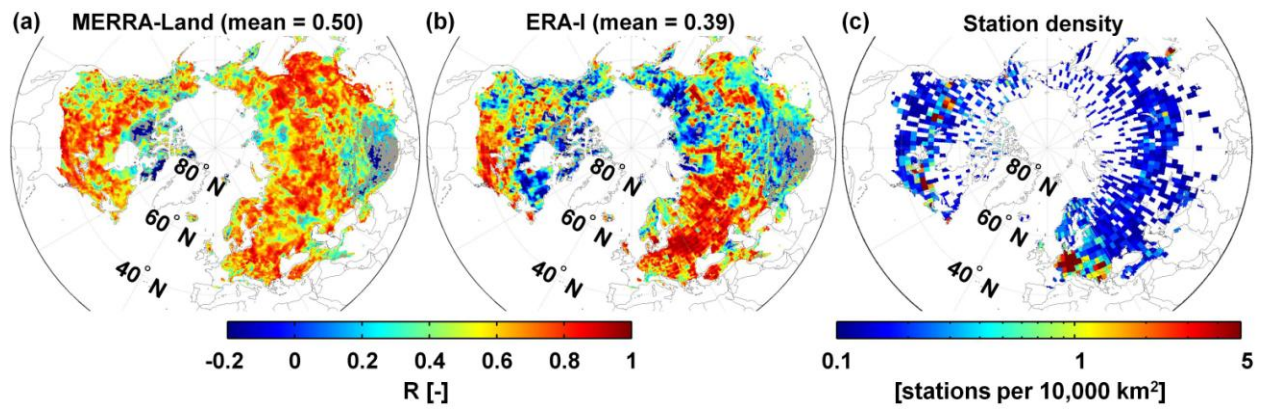
960

961

962

963

964

965

56

Figure 9. Skill (pentad anomaly R) of (a) MERRA-Land and (b) ERA-I snow depth versus CMC

estimates (September 1998 – September 2009). R values that are not statistically different from

zero at the 5% significance level are shown in gray. (c) Maximum density of in situ snow depth

measurements available for CMC snow analysis.